

Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions

Timothy D. Wilson
University of Virginia

Jonathan W. Schooler
University of Pittsburgh

In Study 1, college students' preferences for different brands of strawberry jams were compared with experts' ratings of the jams. Students who analyzed why they felt the way they did agreed less with the experts than students who did not. In Study 2, college students' preferences for college courses were compared with expert opinion. Some students were asked to analyze reasons; others were asked to evaluate all attributes of all courses. Both kinds of introspection caused people to make choices that, compared with control subjects', corresponded less with expert opinion. Analyzing reasons can focus people's attention on nonoptimal criteria, causing them to base their subsequent choices on these criteria. Evaluating multiple attributes can moderate people's judgments, causing them to discriminate less between the different alternatives.

When faced with a difficult decision, people sometimes spend a good deal of time thinking about the advantages and disadvantages of each alternative. At one point or another, most of us have even reached for a sheet of paper and made a list of pluses and minuses, hoping that the best course of action would become clear. Reflection of this kind is generally thought to be beneficial, organizing what might otherwise be a confusing jumble of thoughts and feelings. Benjamin Franklin, for example, relayed the following advice to the British scientist Joseph Priestley about how to make a difficult choice:

My way is to divide half a sheet of paper by a line into two columns, writing over the one Pro, and over the other Con. Then, during three or four days consideration, I put down under the different heads short hints of the different motives, that at different times occur to me, for or against each measure . . . I find at length where the balance lies; and if, after a day or two of further consideration, nothing new that is of importance occurs on either side, I come to a determination accordingly . . . When each [reason] is thus considered, separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to make a rash step. (Quoted in Goodman, 1945, p. 746)

Franklin's advice has been captured, at least in spirit, by many years of research on decision analysis (e.g., Edwards, 1961; Keeney, 1977; Koriat, Lichtenstein, & Fischhoff, 1980; Raiffa, 1968; Slovic, 1982). Though the terms *decision theory* and *decision analysis* describe a myriad of theoretical formulations, an assumption made by most of these approaches is that

decisions are best made deliberately, objectively, and with some reflection. For example, Raiffa (1968) states that

the spirit of decision analysis is divide and conquer: Decompose a complex problem into simpler problems, get your thinking straight in these simpler problems, paste these analyses together with a logical glue, and come out with a program for action for the complex problem (p. 271).

Janis and Mann (1977) go so far as to predict that a "balance sheet" procedure similar to Benjamin Franklin's will become as commonplace among professional and personal decision makers as recording deposits and withdrawals in a bankbook.

Curiously, however, there has been almost no research on the effects of reflection and deliberation on the quality of decision making. One reason for this lack of research is the difficulty of assessing how good any particular decision is. For example, Janis and Mann (1977) arrived at the "somewhat demoralizing" conclusion that there is "no dependable way of objectively assessing the success of a decision" (p. 11). Whereas we agree with Janis and Mann that any one measure of the quality of a decision has its drawbacks, we argue that it is not impossible to evaluate people's decisions, particularly if converging measures are used. The purpose of the present studies was to examine the effects of two different kinds of introspection on decision making. We hypothesized that contrary to conventional wisdom, introspection is not always beneficial and might even be detrimental under some circumstances.

Our studies can be viewed as part of a growing literature on the drawbacks of introspection and rumination. Recent research from a variety of sources casts doubt on the view that introspection is always beneficial. Morrow and Nolan-Hoeksema (1990), for example, found that ruminating about a negative mood was less successful in improving this mood than was engaging in a distracting task. Schooler and Engstler-Schooler (1990) documented a deleterious effect of a different kind of

This research was supported by National Institute of Mental Health Grant MH41841 to Timothy D. Wilson and a grant from the University of Pittsburgh Office of Research to Jonathan W. Schooler. We would like to thank Jack McArdle for his statistical advice.

Correspondence concerning this article should be addressed to Timothy D. Wilson, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903-2477.

reflection: Subjects who verbalized their memory for nonverbal stimuli (such as faces) were less likely than control subjects to recognize these faces on a subsequent recognition test. Most relevant to the present concerns, Wilson and his colleagues found that introspecting about the causes of one's attitudes can have disruptive effects, such as reducing attitude-behavior consistency and changing people's attitudes (Wilson, 1990; Wilson, Dunn, Kraft, & Lisle, 1989; see also Millar & Tesser, 1986a).

Effects of Analyzing Reasons

Forming preferences is akin to riding a bicycle; we can do it easily but cannot easily explain how. Just as automatic behaviors can be disrupted when people analyze and decompose them (Baumeister, 1984; Kimble & Perlmuter, 1970; Langer & Imber, 1979), so can preferences and decisions be disrupted when people reflect about the reasons for their feelings (Wilson, Dunn, Kraft, & Lisle, 1989). We suggest that this can occur as follows. First, people are often unaware of exactly why they feel the way they do about an attitude object. When they reflect about their reasons, they thus focus on explanations that are salient and plausible. The problem is that what seems like a plausible cause and what actually determines people's reactions are not always the same thing (Nisbett & Wilson, 1977). As a result, when asked why they feel the way they do, people focus on attributes that seem like plausible reasons for liking or disliking the stimulus, even if these attributes have no actual effect on their evaluations.

It might seem that people would focus only on attributes of the stimulus that are consistent with their initial attitude, to justify how they feel. That is, even if people do not know why they feel the way they do, and have to construct reasons, they might focus only on factors that could account for their present feelings. Undoubtedly such a justification process can occur. We suggest that under some circumstances, however, people will focus on reasons that imply a different attitude than they held before and will adopt the attitude implied by these reasons. These circumstances are hypothesized to be as follows. First, people often do not have a well-articulated, accessible attitude and thus do not start out with the bias to find only those reasons that are consistent with an initial reaction. They conduct a broader search for reasons, focusing on factors that are plausible and easy to verbalize even if they conflict with how they felt originally.

Even when people's initial attitude is inaccessible, analyzing reasons will not always change their attitude. A cause of people's attitude might be so powerful and obvious that it is difficult to miss when they analyze their reasons. For example, if we knew nothing about a stranger except that he was convicted of child abuse and then were asked why we felt the way we did about him, we would have little difficulty in pinpointing the actual cause of our feelings. Second, even if people miss an important cause of their feelings when they analyze reasons, they will not change their attitudes if the reasons that are salient and plausible are of the same valence as the actual cause. Thus, people might not realize that Attribute A was a major determinant of their reaction and instead might focus on Attribute B. If Attributes A and B imply the same feeling, however, no attitude change will occur.

In sum, we suggest that reflecting about reasons will change people's attitudes when their initial attitude is relatively inaccessible and the reasons that are salient and plausible happen to have a different valence than people's initial attitude. A considerable amount of evidence has been obtained that is consistent with these hypotheses. It is well documented, for example, that when people are asked to think about why they feel the way they do, they sometimes bring to mind reasons that are discrepant from their initial attitude and that they adopt the attitude implied by these reasons (e.g., Millar & Tesser, 1986a; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984; Wilson, Kraft, & Dunn, 1989). In addition, Wilson, Hodges, and Pollack (1990) found that thinking about reasons was most likely to change people's attitudes when their initial attitude was relatively inaccessible.

It has not been clear, however, whether there is any harm done by the attitude change that occurs when people analyze reasons. We suggest that thinking about reasons can alter people's preferences in such a way that they make less optimal choices. In many domains, people have developed an adaptive, functional means of how to weight different information about a stimulus. For example, when evaluating food items with which they are familiar, people have little difficulty deciding which ones they prefer the most. Asking people to think about why they feel that way might focus their attention on attributes that seem like plausible reasons for liking or disliking the items but that in fact have not been heavily weighted before. Similarly, people might dismiss attributes that seem like implausible reasons but that in fact had been weighted heavily before. As a result, they change their mind about how they feel. To the extent that their initial reaction was adaptive and functional, this change might be in a less optimal direction.

Effects of Evaluating Multiple Attributes of Stimuli

A related kind of introspection might also influence people's decisions in disadvantageous ways, but in a different manner. Sometimes, when evaluating a stimulus, people decompose it into many different attributes. For example, potential car buyers sometimes consider a wide array of information about cars—such as their price, safety, repair record, gas mileage, and resale value. There is evidence that evaluating a stimulus on several different dimensions causes people to moderate their evaluations. Linville (1982), for example, asked people to evaluate five different brands of chocolate chip cookies. She asked some subjects to consider six different attributes of the cookies before rating them, such as how sweet they were and the number of chocolate chips they contained. She asked others to consider only two of these attributes. As predicted, those who evaluated six attributes made more moderate evaluations than those who evaluated two attributes: The range and standard deviation of their ratings of the five cookies were significantly smaller.

This moderation effect is most likely to occur when the different attributes people consider are uncorrelated, so that some are positive and some are negative (Judd & Lusk, 1984; Millar & Tesser, 1986b). The more such attributes people consider, the more all the alternatives will seem to have some good and some bad qualities and thus will appear more similar to each other. To our knowledge, no one has examined the effects of consider-

ing multiple attributes of a set of alternatives on the quality of people's decisions. If this type of introspection makes the alternatives more difficult to distinguish from one another, people may be more likely to make a poor choice. And, as noted earlier, to the extent that people's initial preferences (before introspecting) are adaptive, any form of thought that changes people's preferences might lead to less optimal choices.

The present studies examined the effects of analyzing reasons (in Studies 1 and 2) and considering multiple attributes of the alternatives (in Study 2) on people's preferences and choices. We hypothesized that both types of introspection would lead to less optimal decisions, by means of the different mechanisms we have just reviewed. Our measure of the quality of people's preferences and choices was expert opinion. In Study 1, we compared subjects' preferences for different brands of a food item, strawberry jam, with the ratings of these brands by trained sensory experts. We assumed that left to their own devices, people's preferences would correspond reasonably well to the ratings of the experts. We predicted that analyzing the reasons for one's reactions to the jams would change people's preferences. Consistent with our hypothesis that analyzing reasons can produce attitudes that are nonoptimal, we predicted that the preferences of people in the reasons condition would not correspond very well with the experts' ratings of the jams. In Study 2, we examined college students' choices of which courses to take and compared these choices with various kinds of expert opinion about what the best choices were.

Study 1

Method

Subjects

Subjects were 49 undergraduate psychology students (39 men, 10 women) at the University of Washington. They volunteered for a study entitled "Jam Taste Test" in return for course credit and were instructed not to eat anything for 3 hours before the study.

Materials and Ratings of the Experts

We purchased five brands of strawberry jams or preserves that varied in their overall quality, as reported by *Consumer Reports* magazine ("Strawberry Jams," 1985). The *Consumer Reports* rankings were based on the ratings of seven consultants who were trained sensory panelists. These experts rated 16 sensory characteristics (e.g., sweetness, bitterness, aroma) of 45 jams; these ratings were averaged to compute the ranking of each jam (L. Mann, *Consumer Reports* magazine, personal communication, May 15, 1987). The jams we purchased were ranked 1st, 11th, 24th, 32nd, and 44th.

Procedure

Subjects, seen individually, were told that the purpose of the study was to evaluate different kinds of jams under different conditions, as part of a consumer psychology experiment. Experimenter 1 explained that some subjects would taste the jams on crackers, whereas others would taste the jams on plastic spoons. All subjects were told that they had been randomly assigned to the condition in which they would taste the jams on spoons and that after tasting the jams, they would be asked to rate their liking for each one. After receiving these initial instructions and signing a consent form, subjects were randomly assigned to a

control or a reasons analysis condition. Reasons analysis subjects received written instructions asking them to "analyze why you feel the way you do about each" jam, "in order to prepare yourself for your evaluations." They were told that they would be asked to list their reasons for liking or disliking the jams after they tasted them, the purpose of which was to organize their thoughts. They were also told that they would not be asked to hand in their list of reasons. Control subjects did not receive any additional instructions.

All subjects were then asked to sit at a table with five plates, each containing a plastic spoon with approximately $\frac{1}{2}$ teaspoon (3.3 ml) of strawberry jam. The jams were labeled with a letter from A to E and were presented in one random order. Experimenter 1 left the room, during which time the subjects tasted each of the five jams.

Version 1. The first five subjects in each condition followed a slightly different procedure than did those who followed. The initial subjects in the reasons analysis condition completed the reasons questionnaire while they tasted the five jams; that is, they tasted Jam 1, listed their reasons for liking or disliking Jam 1, tasted Jam 2, listed their reasons for liking or disliking Jam 2, and so on. The experimenter reiterated that the purpose of this questionnaire was to organize the subjects' thoughts and that they would not be asked to hand it in. When she returned, she picked up the reasons questionnaire, explained that it would not be needed anymore, and deposited it in a trash can. The initial subjects in the control condition tasted all five jams and then rated each one, without filling out any questionnaires.

Version 2. To equalize the amount of time subjects spent on the tasting part of the study, subsequent subjects followed a slightly different procedure. All subjects tasted the jams without filling out any questionnaires and then were given a questionnaire to fill out when the experimenter returned. Subjects in the reasons condition received the reasons questionnaire. As in Version 1, they were told that they would not hand in this questionnaire, and the experimenter deposited it in the trash when she returned. Subjects in the control condition received a filler questionnaire instructing them to list reasons why they chose their major. The experimenter also left the room while control subjects completed this questionnaire. She collected the questionnaire when she returned.

The remainder of the experiment was identical for all subjects. Experimenter 1 introduced subjects to Experimenter 2, who was unaware of whether they had analyzed reasons. Experimenter 2 gave subjects a questionnaire on which to evaluate the jams, which consisted of a 9-point scale ranging from *disliked*(1) to *liked*(9) for each jam. Subjects were instructed to complete the questionnaire and to place it through a slot in a covered box, to maintain anonymity. Experimenter 2 left the room while subjects made their ratings. He fully debriefed subjects when he returned.

Results

We predicted that asking subjects to think about reasons would change their evaluations of the jams. Consistent with this prediction, a multivariate analysis on the mean ratings of the five jams found a significant effect of the reasons analysis manipulation, $F(5, 43) = 3.09, p = .02$. Individual *t* tests were significant on two of the jams, as seen in Table 1. We also predicted that analyzing reasons would produce preferences that were, in some sense, nonoptimal. To test this prediction, we computed the Spearman rank-order correlation between each subject's ratings of the five jams and the rank ordering of the jams by the *Consumer Reports* taste experts (for all analyses, these within-subject correlations were converted to *z* scores by means of Fisher's *r*-to-*z* transformation; the means reported here have been converted back to correlation coefficients). The

Table 1
Study 1: Mean Liking Ratings for the Five Jams

Condition	Jam 1	Jam 2	Jam 3	Jam 4	Jam 5
Control					
M	6.52	7.64	6.12	2.72	4.68
SD	2.22	1.66	2.05	2.26	2.66
Reasons					
M	4.54	6.25	5.42	2.88	4.92
SD	2.00	2.38	2.70	2.13	2.89
t	3.27	2.38	1.03	-.25	-.30
p	.002	.02	.31	.81	.77

Note. The jams are listed in order of their rankings by the *Consumer Reports* experts; Jam 1 was the highest ranked jam, Jam 2 was the second highest, and so on. The liking ratings were made on 9-point scales that ranged from *disliked* (1) to *liked* (9).

mean correlation in the control condition was .55, reflecting a fair amount of agreement with the taste experts. As predicted, the mean correlation in the reasons condition was significantly lower ($M = .11$), $t(47) = 2.53$, $p = .02$.¹ The mean correlation in the control condition was significantly higher than zero, $t(24) = 4.27$, $p = .0003$, whereas the mean correlation in the reasons condition was not, $t(23) = .80$, $p = .43$.

We noted earlier that some kinds of introspection cause people to moderate their evaluations. We have not found this to be the case with analyzing reasons in previous studies (e.g., Wilson, Lisle, & Schooler, 1990). Nor does analyzing reasons reduce people's confidence in their attitudes (Wilson, Dunn, Kraft, & Lisle, 1989). Nonetheless, it is important to see if in the present study, asking people to explain their preferences led to moderation. If so, this reduced variability in people's ratings might account for the lower correlation between their ratings and the opinions of the *Consumer Reports* experts. Though the mean ratings of the jams displayed in Table 1 seem to support this interpretation (i.e., the range in ratings of the five jams was lower in the reasons condition), it is more appropriate to test this possibility on a within-subject basis.² We computed the range between each subject's highest and lowest rating of the jams, as well as the standard deviation of each subject's ratings. On average, these values were quite similar in both the reasons and control conditions, $t(47) < .39$, $p > .71$. Thus, there was no evidence that analyzing reasons caused people to evaluate the jams more similarly than did control subjects.

Instead, people seemed to have come up with reasons that conflicted with the experts' ratings and adopted the attitude implied by these reasons. Support for this interpretation comes from analyses of the reasons people wrote down in the reasons condition. Subjects' responses were first divided into individual reasons by a research assistant and then put into different categories of reasons for liking or disliking the jams. (Another research assistant coded a subset of the questionnaires and agreed with the first assistant's initial divisions into reasons 95% of the time and agreed with her placement of the reasons into individual categories 97% of the time.) Subjects gave an average of 2.93 reasons per jam. These reasons concerned some aspect of their taste (e.g., sweetness, tartness, fruitiness, 52%), texture (e.g.,

thickness, chunkiness, ease of spreading, 35%), appearance (e.g., color, how fresh they looked, 8%), smell (1%), naturalness or artificiality of the ingredients (1%), and miscellaneous (3%). Two research assistants also coded, on a 7-point scale, how much liking for each jam was expressed in subjects' reasons (reliability $r = .97$). Consistent with our hypothesis that the reasons people came up with would not match expert opinion, this index did not correlate significantly with the experts' ratings of the jams ($M = .25$), $t(23) = 1.74$, $p > .09$. Consistent with our hypothesis that people would base their attitude on the reasons they listed, this index correlated very highly with subjects' subsequent ratings of the jams (mean within-subject correlation = .92), $t(23) = 8.60$, $p < .0001$.

A closer look at how analyzing reasons changed people's attitudes is illuminating. In some of our previous studies, people who analyzed reasons changed their attitudes in the same direction, possibly because similar attributes of the stimuli became salient when people analyzed reasons, and people held similar causal theories about how these attributes affected their judgments (e.g., Wilson et al., 1984). In other studies, the attitude change was more idiosyncratic (e.g., Wilson, Kraft, & Dunn, 1989), which can occur for at least two reasons. First, for some stimuli, the attributes that become salient might differ from person to person. For example, when asked why they feel the way they do about a political candidate, people draw on different knowledge bases. The fact that is most salient to one person (e.g., that the candidate is antiabortion) may be completely unknown to another. Second, even if the same fact, such as the candidate's stance on abortion, is available to everyone, it may be evaluated quite differently by different people, leading to attitude change in different directions.

The fact that there were significant differences between conditions on ratings of two of the jams (see Table 1) indicates that at least some of the change in the present study was in a common direction: Subjects who analyzed reasons became more negative, on average, toward Jams 1 and 2. However, other changes may have occurred in idiosyncratic directions, so that some people who analyzed reasons became more positive, whereas others became more negative. To test this possibility, we correlated each subject's ratings of the five jams with the ratings of every other subject in his or her condition and then

¹ Initial analyses revealed that the effects of analyzing reasons did not differ according to which version of the procedure was used. Subjects in both conditions who followed the initial procedure—in which the jams were rated right after tasting them, without an intervening questionnaire—had higher correlations between their ratings of the jams and the *Consumer Reports* experts' ratings of the jams, as indicated by a significant main effect of version ($p = .02$). The difference in correlations between the reasons and control conditions, however, was in the same direction in both versions, and the Reasons \times Version interaction was nonsignificant ($p = .60$). Initial analyses also revealed that there were no significant effects of gender; thus subsequent analyses were collapsed across this variable.

² For example, consider two hypothetical subjects in the reasons condition, one of whom gave ratings of 9, 7, 5, 3, and 1 to the five jams, the other of whom gave ratings of 1, 3, 5, 7, and 9. The mean of these two subjects' ratings would be 5 for every jam, making it appear as though they were not discriminating between the jams, when in fact they were making very strong discriminations.

averaged these correlations, using Fisher's r -to- z -to- r transformation. The average correlation in the control condition was .55, indicating a fair amount of consensus about how likable the jams were. If subjects in the reasons condition changed their attitudes in a common direction, then their ratings should have correlated as highly, or possibly even higher, with other subjects in this condition. If these subjects changed their attitudes in idiosyncratic directions, then there should have been less consensus in the reasons condition. Supporting this latter possibility, the mean intercorrelation in the reasons condition was significantly lower than in the control condition ($M = .18$), $t(47) = 4.38$, $p < .0001$.³

Discussion

Left to their own devices, control subjects formed preferences for strawberry jams that corresponded well to the ratings of trained sensory experts. Subjects asked to think about why they liked or disliked the jams brought to mind reasons that did not correspond very well with the experts' ratings. They then seem to have based their preferences on these reasons (i.e., the correlation between the attitude implied by their reasons and their subsequent preferences was extremely high). As a result, their preferences did not correspond as well with expert opinion. No evidence was found for the possibility that analyzing reasons moderated subjects' judgments. Instead it changed people's minds about how they felt, presumably because certain aspects of the jams that were not central to their initial evaluations were weighted more heavily (e.g., their chunkiness or tartness).

It might be argued that there should have been a greater correspondence between the experts and subjects who analyzed reasons, because both sets of people made their ratings in an analytical frame of mind. The ratings made by the two groups, however, differed in important ways. First, the experts were provided in advance with a list of 16 criteria on which to evaluate the jams (L. Mann, *Consumer Reports* magazine, personal communication, May 15, 1987). In contrast, our reasons subjects had to decide for themselves which criteria to use, increasing the probability that they would focus on a few attributes that were salient and plausible as causes of their preferences. Second, the experts were trained sensory panelists with a good deal of experience in tasting food items. Wilson, Kraft, and Dunn (1989) found that people who are knowledgeable about the attitude object are unaffected by analyzing their reasons. Thus, even if the experts evaluated the jams analytically, we would expect their ratings to differ from the subjects in our reasons condition, who were not experts.

It might also be argued that the different attitudes reported by subjects in the reasons condition were due to demand characteristics. Though we went to some length to convince these subjects that no one would see their reasons, they still might have believed we would compare their attitude responses with their reasons, and thus they might have purposely exaggerated the similarity of their attitudes to their reasons because of concerns about consistency. Note, however, that even if this interpretation were true, it would not explain why the reasons generated by subjects implied an attitude that was different from those held by control subjects and the *Consumer Reports* experts.

One way to rule out a demand characteristics explanation more definitively would be to allow people to choose one of the attitude objects for their own personal use. For example, suppose we had told subjects in Study 1 that they could choose one of the jams to take home and had set up the study in such a way that no one would know which brand subjects chose. If subjects in the reasons condition acted on their reported attitudes—that is, if they chose jams that they had rated highly—it would seem that they had genuinely changed their attitudes, rather than simply reporting a new attitude to please the experimenter. Though we did not follow such a procedure in Study 1, we did in two studies by Wilson et al. (1990). For example, in one study, subjects examined five art posters and chose one to take home. The results were inconsistent with a demand characteristics explanation: Subjects who analyzed reasons chose different posters, even though they believed that the experimenter would not know which one they chose.

The Wilson et al. (1990) studies addressed another possible concern with Study 1: the use of expert opinion as our criterion of decision quality. It might be argued that even though subjects in the reasons condition formed preferences that were at variance with the experts, there was no cost in doing so. As long as people like a particular kind of jam, what difference does it make that experts disagree with them? We suggest it can make a difference, because the attitude change caused by analyzing reasons is often temporary. Over time, people probably revert to the weighting schemes they habitually use. If they made a choice on the basis of a different weighting scheme, they might come to regret this choice. To test this prediction, Wilson et al. (1990) contacted subjects a few weeks after they had been in the study, and asked them how satisfied they were with the poster they had chosen. As predicted, subjects who analyzed reasons expressed significantly less satisfaction with their choice of poster. Thus, analyzing reasons has been shown to reduce the quality of preferences in two different ways: It can lower the correspondence between these preferences and expert opinion, and it can cause people to make decisions they later regret.

Study 2 attempted to extend these findings in a number of respects. First, it was a field experiment that examined a real-life decision of some importance to college students: their choice of which courses to take the following semester. Students were presented with detailed information about all of the soph-

³ Two points should be made about these mean intercorrelations: one statistical and one conceptual. First, the lowered consensus in the reasons condition might show that people's evaluations became more random—that is, by becoming unsure of how they felt, subjects' ratings contained more "error," and thus were not as correlated with each other. Though we cannot completely rule out this interpretation, the fact that analyzing reasons did not reduce the range in subjects' ratings and the fact that in previous studies, analyzing reasons has not made people less confident in their evaluations, reduces its plausibility (see Wilson, Dunn, Kraft, & Lisle, 1989). Second, note that to avoid the problem of lack of independence of the intercorrelations (e.g., there were 300 intercorrelations among the 25 subjects in the control condition), the t test was computed on the mean of each subject's intercorrelations with every other subject in his or her condition, so that there was one data point for each subject.

omore-level psychology courses being offered the next semester, and we examined their ratings of each course and whether they actually registered for the different courses. As in Study 1, we included a measure of expert opinion of the desirability of the alternatives. The "experts" were students who had previously taken the courses. We predicted that subjects in the control conditions would be most likely to choose courses recommended by these experts; that is, they should be most likely to register for the courses that had received the highest course evaluations. Subjects who analyzed reasons, however, might change the criteria they used to make their decision and thus be less likely to sign up for the highly rated ones.

Second, as discussed in the Introduction, we examined the effects of another form of introspection, in addition to analyzing reasons. Some subjects were asked to consider how every attribute of every course (e.g., the topic matter, the time it met) influenced their preferences. We hypothesized that this form of introspection would moderate subjects' ratings of the courses, by making them more cognizant of the fact that every course had pluses and minuses (Linville, 1982). We also hypothesized that this form of introspection might confuse subjects about which information was the most important, causing them to assign more equal weights to the different information. This change in subjects' weighting scheme was also expected to change their decisions about which courses to take, possibly in a nonoptimal direction.

Third, we included a long-term measure of subjects' behavior: the courses they were enrolled in at the end of the following semester. Subjects had the opportunity to add and drop courses at the beginning of the semester; thus, even if our manipulations influenced their initial decision of which courses to take, they could revise these decisions later. Whether the manipulation would influence subjects' long-term behavior was an open question. On the one hand, we have argued that the attitude change caused by analyzing reasons is relatively temporary and will not influence long-term behavior. Consistent with this view, Wilson et al. (1984, Study 3) found that analyzing reasons did not influence dating couple's decision about whether to break up several months after the study was completed. On the other hand, if analyzing reasons changes subjects' decisions about the courses for which they register, they might experience a certain amount of inertia, so that they remain in these courses, even if they change their mind at a later point. Furthermore, Millar and Tesser (1986a, 1989) found that analyzing reasons highlights the cognitive component of attitudes and that these cognitively based attitudes will determine behaviors that are more cognitively based than affectively based. Given that the decision of whether to take a college course has a large cognitive component (e.g., whether it will advance one's career goals), the attitude change that results from analyzing reasons might cause long-term changes in behavior.⁴

Fourth, to test more directly the hypothesis that people who analyze reasons change the criteria they use to make decisions, we included some additional dependent measures assessing the criteria subjects used, and we compared these criteria with another kind of expert opinion: ratings by faculty members in psychology of the criteria students ought to use when choosing courses. We predicted that the criteria used by control subjects would correspond at least somewhat to the criteria faculty

members said students ought to use but that there would be less of a correspondence in the reasons condition. This would be consistent with our hypothesis that analyzing reasons can cause people to alter the criteria they use in nonoptimal ways.

Study 2

Method

Subjects

Two hundred and forty-three introductory psychology students at the University of Virginia volunteered for a study entitled "Choosing College Courses." The sign-up sheet indicated that participants would receive detailed information about all of the 200-level courses being offered by the psychology department the following semester (i.e., sophomore-level courses) and that only students who were considering taking one of these courses should volunteer for the study. Thirteen students were eliminated from the analyses for the following reasons: One participated in the study twice, 2 reported that they would not be enrolled in college the next semester, and 10 reported that they had already registered for classes, which was one of the major dependent variables. Other subjects failed to complete some of the individual questions and were eliminated from the analyses of these measures. Subjects received course credit for their participation.

Procedure

Subjects were run in large groups in the first 2 days of the preregistration period, when students register for the classes they want to take the following semester. Subjects received written instructions indicating that the purpose of the study was both to provide people with more information than they would ordinarily receive about 200-level psychology courses and to "look at some issues in decision making of interest to psychologists, such as how people make decisions between alternatives." They were given a packet of materials and told to go

⁴ We should address some possible ethical objections to Study 2. It might be argued that it was unfair to ask subjects to reflect about their decision of which courses to take, given our hypothesis that it would change the courses for which they preregistered and possibly even change the courses they actually took the following semester. We struggled with this issue before conducting the study and discussed it with several colleagues. In the end, we decided that the potential knowledge gained—discovering some detrimental effects of introspection—outweighed the possible harmful effects on the participants. It would have been unacceptable to give subjects misinformation about the courses—for example, telling them that a course was highly rated by students when in fact it was not. However, we gave all subjects accurate information and then asked some of them to reflect more than they might ordinarily do when forming their preferences. According to the predominant theories of decision making (e.g., Janis & Mann, 1977), asking people to be more reflective about their choices should have beneficial effects. Probably thousands of decision analysts, counselors, and academic advisers urge people to make decisions in ways similar to subjects in our reasons and rate all conditions. Given that the effects of our manipulations were predicted to be relatively benign (altering the psychology courses for which subjects preregistered and possibly altering the courses they took the following semester), we felt it was worth testing the wisdom of such advice. We did not, of course, make this decision alone. The study was approved by a Human Subjects Committee.

through it page by page without looking ahead, though they could look back at any point. After filling out some demographic information, they received descriptions of the nine 200-level psychology classes.

Course descriptions. Each course description included the name of the professor teaching the course, when and where it would meet, the required and recommended prerequisites for the course, the requirements for the psychology major satisfied by the course, whether a term paper was required, the format of the course (lecture or discussion), evaluations of the course by students who took the course the last time it was taught by the same professor, whether there was a required or optional discussion section, a description of the course contents, and a list of the books to be used. The course evaluations included a frequency distribution of the responses to two ratings, the overall teaching effectiveness of the instructor and the intellectual stimulation of the course, as well as the mean response to these two questions. Most, though not all, of this information was available for all nine courses. For example, one course was being taught by a new instructor—thus course evaluations were not available—and the format of one course was unknown. The course descriptions were presented in one of two counterbalanced orders.

Experimental conditions. Subjects were randomly assigned to one of three experimental conditions within each group session. In the rate all information condition (hereafter referred to as rate all), subjects were asked to stop and think about each piece of information about every course and then to rate the extent to which it made them more or less likely to take the course. Underneath each item, subjects were reminded to "stop and think about this piece of information," after which they rated it on a 9-point scale ranging from *makes me much less likely to take it* (1) to *makes me much more likely to take it* (9). Subjects in the reasons condition were instructed to think about why they might want or not want to take a course as they read the course descriptions. They were told that they would be asked to write down their reasons and were asked to prepare themselves by "analyzing why you feel the way you do about each course." After reading the course descriptions (without making any ratings of the information), these subjects did in fact write down their reasons for each of the nine courses. They were told that the purpose of this was to organize their thoughts and that their responses would be completely anonymous. They were also reminded that they could refer back to the course descriptions if they wanted. Subjects in the control condition were instructed to read the information about the nine courses carefully, after which they received a filler questionnaire that asked their opinion of some university issues (e.g., what they thought about the advising and honor systems) and their leisure-time activities.⁵

Dependent Measures

All subjects rated the likelihood that they would take each course on a scale ranging from *definitely will not take this course* (1) to *definitely will take this course* (9). If they had already taken a course, they were asked to indicate this and to not complete the rating scale. The courses were rated in the same order as they were presented in the course description packet. Subjects next rated each type of information they had received about the courses (e.g., the course evaluations, the course content), as well as two additional pieces of information (what they had heard about the courses from other students or professors and how interested they were in the topic), according to how much it influenced their decision about which courses to take. These ratings were made on scales ranging from *did not influence me at all* (1) to *influenced me a great deal* (9). The information about the courses was rated in one of two counterbalanced orders.

At this point, subjects handed in their packets and were given, unexpectedly, a recall questionnaire. They were asked to recall as much information about the courses as they could and to write it down in

designated spaces for each course. Their responses were later coded by a research assistant who was unaware of the subjects' condition. She assigned subjects a 1 for each piece of information recalled correctly, a 0 for each piece not recalled, and a -1 for each piece recalled incorrectly. One of the authors also coded the recall questionnaires of 7 subjects; his codings agreed with the research assistant's 94% of the time.

After completing the recall measure, subjects were asked to sign a release form giving us permission to examine the registrar's records so that we could record the courses for which they actually registered. All subjects agreed to sign this form. They were then given a written explanation of the study that explained it in general terms; that is, that the study was concerned with the kinds of information people use when deciding what courses to take. Neither the hypotheses nor the different conditions of the study were discussed. At the end of the following semester, all subjects were sent a complete written description of the purpose of the study.

Expert Opinion on the Criteria for Choosing Courses

A questionnaire was distributed to the 34 faculty members in psychology in residence at the University of Virginia. They were given a description of the 10 pieces of information subjects had received about the psychology courses (e.g., "whether or not a term paper is required"), as well as the two other pieces of information that subjects had rated (what the student had heard about the courses from other students or professors and how interested the student was in the topic), in one of two counterbalanced orders. The faculty rated how much students should use each piece of information "to make sure they make the best decision they can" about which 200-level psychology course to take. These ratings were made on scales ranging from *should be given very little weight* (1) to *should be weighted very heavily* (9). A total of 18 (53%) of the faculty completed the questionnaire.

⁵ The inclusion of the filler questionnaire in the control condition solved one problem but possibly created another. The problem it solved was controlling for the amount of time that elapsed between the examination of the course descriptions and the completion of the dependent variables in the reasons condition. It also, however, made the control and reasons conditions different in the amount of time spent thinking about unrelated matters between the examination of the courses and the dependent measures. That is, subjects in the reasons condition read the descriptions, spent several minutes thinking about why they felt the way they did about the courses, and then rated the courses. Control subjects spent several minutes thinking about unrelated matters after reading the course descriptions, which might have adversely affected their memory for the courses. To correct this problem, two versions of the control condition were run: one in which subjects completed the filler questionnaire between reading the descriptions and completing the dependent measures, to equalize the delay between these activities, and one in which subjects completed the dependent measures immediately after reading the descriptions so that they would not be distracted by thinking about unrelated matters before completing the dependent measures. As it happened, the presence or absence of the delay in the control group produced very few significant differences on the dependent measures. The only difference was that subjects who had no delay between the course descriptions and the dependent measures reported that they were significantly less likely to take two of the nine courses. Because there were no other differences on any other dependent measure (including the actual registration and enrollment figures and the recall data), the data from the two versions of the control condition were combined in all analyses reported later.

Results

Initial analyses revealed that neither the order in which the courses were presented, the order in which subjects rated how much the information about the courses influenced their likelihood of taking them, nor subjects' gender interacted significantly with the independent variables. There were a few significant main effects of gender and course order; for example, women recalled more information about the courses than did men, and the order in which the courses were presented had a significant effect on subjects' ratings of how likely they were to take some of the courses. Because the distributions of men and women and of people who received the courses in each order were nearly identical in each condition, however, we collapsed across gender and order in all subsequent analyses.

Recall for and Ratings of Influence of the Course Information

We predicted that the two introspection manipulations would alter the way subjects weighted the different information about the courses. To test this, we examined their recall for the information and their ratings of how much each type of information had influenced their decisions. We would certainly not argue that these measures were perfectly correlated with the weights subjects actually assigned to the different criteria. As one of us has noted elsewhere, subjects' causal reports are often inaccurate (Nisbett & Wilson, 1977). It is also well known that recall is often uncorrelated with people's weighting schemes (Hastie & Park, 1986). Few would argue, however, that such measures were orthogonal to the weights people used. Thus, relative differences in reported influence and recall between different conditions can be taken as rough indicators of what subjects in those conditions found important about the courses (Anderson & Pichert, 1978).

Recall. Interestingly, the total amount of information subjects recalled did not differ across the three conditions, $F(2, 226) < 1$. There were, however, differences in the kinds of information subjects recalled. Subjects' recall scores were averaged across the nine courses and analyzed in a 3 (introspection condition) \times 10 (type of information, e.g., when the course met, whether a term paper was required) analysis of variance (ANOVA), with the last factor treated as a repeated measure. There was a very strong effect for type of information, $F(10, 217) = 59.53, p < .001$, reflecting the fact that subjects were more likely to recall some kinds of information about the courses than they were others. More interestingly, there was also a significant Condition \times Type of Information interaction, $F(20, 434) = 2.53, p < .001$, indicating that the kinds of information subjects were most likely to remember differed by condition.

How well did subjects' recall correspond to the opinion of faculty as to how much people should weight each piece of information? We predicted that subjects in the control condition would do a reasonably good job of attending to the information that was important about the courses, whereas the introspection manipulations might disrupt this process. To test this prediction, we averaged subjects' recall for the three pieces of information faculty rated as most important (who was teaching the class, the course content, and the prerequisites for the

class) and subjects' recall for the three pieces of information faculty rated as least important (when the class met, whether there was a required term paper, and whether the course had a discussion section). As seen in Table 2, control subjects recalled more of the "important" than "unimportant" information, $F(1, 226) = 10.09, p < .01$. As predicted, this was not the case in the two introspection conditions. Subjects in the reasons condition were no more likely to recall important than unimportant information, and subjects in the rate all condition actually recalled more of the unimportant information, $F(1, 226) = 3.46, p = .06$. These results were reflected by a significant Condition \times Importance of Information interaction, $F(2, 226) = 8.28, p < .001$. This interaction was also significant when the control condition was compared with the reasons condition alone, $F(1, 226) = 5.25, p < .05$, and with the rate all condition alone, $F(1, 226) = 12.69, p < .001$.

Ratings of influence of the course information. Subjects rated how much each of the 10 pieces of information about the courses influenced how likely they were to take them, as well as the influence of 2 additional items: what they had heard about the course from others and how interested they were in the topic of the course. A 3 (condition) \times 12 (information type) between/within ANOVA revealed a significant main effect for condition, $F(2, 223) = 8.46, p < .001$, reflecting the fact that subjects in the rate all condition ($M = 5.78$) thought that all of the information had influenced them more than did subjects in the control and reasons conditions ($M_s = 5.17$ and 5.26 , respectively). The ANOVA also yielded a significant Condition \times Information Type interaction, $F(22, 426) = 2.81, p < .001$, indicating that the manipulations influenced what kinds of information subjects thought influenced them.

As seen in Table 2, control subjects reported that the important information influenced them more than did the unimportant information, $F(1, 223) = 50.42, p < .001$. In contrast, subjects in the rate all condition reported that the two types of information had influenced them about equally, $F(1, 223) < 1$. Unexpectedly, subjects in the reasons condition responded similarly to control subjects. A 3 (condition) \times 2 (importance of information) between/within ANOVA revealed a highly significant interaction, $F(2, 223) = 9.20, p < .001$. This interaction

Table 2
Recall for and Reported Influence of the Course Information as a Function of the Importance Attributed to These Items by Faculty

Variable	Condition		
	Control	Reasons	Rate all
Recall			
Recall for 3 highest items	0.23	0.19	0.16
Recall for 3 lowest items	0.14	0.19	0.21
Ratings of influence			
Ratings of 3 highest items	6.41	6.47	6.26
Ratings of 3 lowest items	4.73	5.11	6.32

Note. The higher the number, the more subjects recalled the information or thought the information influenced their decision of what courses to take.

was also significant when considering the control and rate all conditions alone, $F(1, 223) = 30.91, p < .001$. It was not significant when the control condition was compared with the reasons condition, $F(1, 223) = 1.06$.⁶

We predicted that the rate all manipulation might confuse people about which attributes of the courses were most important, causing them to assign more equal weights to the different information. One piece of evidence for this prediction was that as just seen, subjects in the rate all condition rated all of the information, on average, as more influential than subjects in the other two conditions. Another was that the mean, within-subject range in subjects' ratings of the influence of the information was significantly smaller in the rate all condition ($M = 6.78$) than in the control and reasons conditions ($M_s = 7.35$ and 7.47 , respectively), $t_{s(224)} > 3.31, ps < .001$. An identical pattern of results was found in an analysis of the within-subject standard deviations of the ratings of the course information.

Reported Likelihood of Taking Each Course

We expected that people instructed to reflect about their decision (i.e., those in the reasons and rate all conditions) would change their minds about which courses were the most desirable and that this change would be in a nonoptimal direction. To test this prediction, we computed the mean of subjects' reported likelihood of taking the five courses that had received the highest course evaluations by students who had taken the classes and the mean ratings of the three that had received the lowest ratings plus one for which no ratings were available (the results are nearly identical if this latter course is eliminated from the analyses). These means were analyzed with a 3 (condition) \times 2 (course evaluation) between/within ANOVA.

The main effect for condition was not significant, $F(2, 199) = 1.88, p > .15$, indicating that subjects' condition did not influence their reported likelihood of taking psychology courses. The main effect for course evaluation was highly significant, $F(2, 199) = 195.61, p < .001$, reflecting the fact that subjects in all conditions preferred the highly rated courses to the poorly rated courses (see Table 3). Most relevant to our hypotheses, the Condition \times Course Evaluation interaction was also significant, $F(2, 199) = 10.80, p < .001$. As predicted, subjects in the control condition showed more of a preference for highly rated courses than for poorly rated courses than subjects in the rate all condition (see Table 3). Considering these two conditions alone, the Condition \times Course interaction was significant, $F(1, 199) = 14.25, p < .001$. Unexpectedly, there were no significant differ-

ences in the reports of subjects in the control versus reasons condition.

To see if subjects in the rate all condition moderated their ratings of the courses, we examined the range of each subjects' ratings of the nine courses. As predicted, the average range was significantly smaller in the rate all condition ($M = 5.19$) than in the control condition ($M = 6.01$), $t(224) = 3.18, p < .001$. The mean in the reasons condition was actually larger than in the control condition ($M = 6.53$), $t(224) = 1.95, p = .05$. An identical pattern of results was found in an analysis of the within-subject standard deviations of the ratings of the courses. Finally, we examined the intercorrelations between subjects' ratings within each condition, as we did in Study 1. The mean intercorrelations in the control and reasons conditions were very similar ($M_s = .24$ and $.23$, respectively). Both of these means were significantly higher than the mean in the rate all condition ($M = .16$), $t_{s(221)} > 2.31, ps < .02$. The lower agreement in the rate all condition may be a result of the fact that there was less variation in these subjects' ratings—that is, the restricted variance in their ratings placed limits on the magnitude of the intercorrelations.

Course Preregistration and Enrollment

In the few days after our study, all the participants registered for the courses they wanted to take the next semester. We obtained the preregistration records for the nine psychology courses and assigned subjects a 1 if they had preregistered for a course, a 0 if they had not, and a missing value if they had already taken the course. We also analyzed the actual course enrollment data at the conclusion of the following semester, to see if any differences found in the preregistration data persisted, even after students had had the option to add and drop courses. These data were coded in an identical fashion to the preregistration data.

Preregistration for courses. As predicted, the two introspection manipulations influenced the kind of courses for which subjects preregistered. As seen in Table 4, subjects in the introspection conditions (especially those who analyzed reasons) were less likely than control subjects to take the highly rated courses but about equally likely to take the poorly rated courses. The number of courses of each type that subjects registered for were analyzed in a 3 (condition) \times 2 (course evaluation) between/within ANOVA, which yielded the predicted Condition \times Course Evaluation interaction, $F(2, 206) = 6.40, p = .002$. This interaction was significant when the control and reasons conditions were considered alone, $F(1, 206) = 12.58, p < .001$, and when the control and rate all conditions were considered alone, $F(1, 206) = 4.12, p < .05$.

It can be seen by the low averages in Table 4 that the modal response in all conditions was not to take any of the nine psychology courses. Despite our request that people only partici-

Table 3
Ratings of Likelihood of Taking the Courses

Evaluation of course	Condition		
	Control	Reasons	Rate all
Highly rated	4.77	4.55	4.45
Poorly rated	3.18	2.85	3.74

Note. The higher the number, the greater the reported likelihood that students would take the class.

⁶ Subjects' ratings of the influence of and their recall for the course information were analyzed in several alternative ways. For example, we computed the within-subject correlations between subjects' recall and the faculty members' ratings of importance and then averaged these correlations across conditions. The results of these and other analyses were very similar to those reported in the text.

Table 4
Courses Preregistered for and Actually Taken

Variable	Condition		
	Control	Reasons	Rate all
Preregistration			
Highly rated courses	.41	.15	.21
Poorly rated courses	.04	.10	.01
Actual enrollment			
Highly rated courses	.37	.21	.24
Poorly rated courses	.03	.08	.03

Note. Subjects were assigned a 1 if they registered for or actually took a course and a 0 if they did not register or take a course.

pate in the study if they were considering taking a 200-level psychology course, many subjects opted not to take any. This created a bit of a statistical anomaly, in that the people who did not take any psychology classes lowered the variance and increased the sample size, thereby increasing the power of the significance tests. To avoid this problem, a 3 (condition) \times 2 (course evaluation) chi-square analysis was performed after eliminating those students who did not register for any of the nine courses. This analysis was also significant, $\chi^2(2, N = 74) = 8.25$, $p = .02$, reinforcing the conclusion that the manipulations influenced the courses for which subjects registered.

Enrollment at the conclusion of the following semester. We did not make firm predictions about whether the effects of the introspection manipulations on people's choice of courses would persist over the long run. To see if they did, we analyzed the course enrollment data at the conclusion of the semester in the same manner as the preregistration data. The results were similar, though not as strong (see Table 4). The interaction effect in a 3 (condition) \times 2 (course evaluation) ANOVA was significant, $F(2, 206) = 3.05$, $p = .05$. This interaction was significant when the control condition was compared only with the reasons condition, $F(1, 206) = 5.90$, $p < .05$, but not with the rate all condition, $F(1, 206) = 2.37$, $p = .13$. The chi-square on only those subjects enrolled in at least one course was not significant, $\chi^2(2, N = 74) = 2.84$, $p = .24$.

To test more definitively whether the effect of the manipulations had weakened over time, the preregistration and final enrollment data were entered into a 3 (condition) \times 2 (course evaluation) \times 2 (time of measurement: registration vs. final enrollment) ANOVA; the last two factors were treated as repeated measures. The Condition \times Course Evaluation interaction was highly significant, $F(2, 206) = 5.31$, $p = .006$, reflecting the fact that at both times of measurement, subjects in the introspection conditions were less likely to take the highly rated courses but about equally likely to take the poorly rated courses. The Condition \times Course Evaluation \times Time of Measurement interaction was not significant, $F(2, 206) = 1.13$, $p = .32$, indicating that the attenuation of the Condition \times Course interaction over time was not reliable.

Other Analyses

Coding of reasons given in the reasons condition. The reasons protocols were coded as described in Study 1, with similar levels

of reliability. Subjects gave an average of 2.06 reasons for liking or disliking each course. The most frequently mentioned reasons were interest in the material (33%), the course evaluations (23%), the course content (13%), whether a term paper was required (7%), and when the course met (6%). The reasons were also coded according to how much liking for each course they conveyed (reliability $r = .98$). The average within-subject correlation between these ratings and subjects' ratings of how likely they were to take each course was .70, $t(63) = 10.93$, $p < .0001$.

Other factors potentially influencing course selection. Some preference is given to upper-level students and majors when they enroll for psychology courses. This could not have accounted for the present results, however, because the number of such students was randomly distributed across conditions, $\chi^2(6, N = 229) = 4.49$, $p = .61$, for upper-level students; $\chi^2(2, N = 230) = 1.07$, $p = .58$, for majors.

Grades obtained in the psychology courses. The grades received by those subjects who took one or more of the nine psychology courses were obtained from the final grade sheets. There were no significant differences between conditions in these grades. The means for the control, reasons, and rate all conditions, on a 5-point scale ranging from A (4) to F (0), were 2.82, 2.78, and 3.20, respectively.

Discussion

We predicted that subjects who introspected about their decision about which courses to take would change the way they evaluated the courses, causing them to make less optimal choices. The results in the rate all condition, in which subjects rated each piece of information about every course according to how it influenced their decision, were entirely consistent with this prediction. These subjects' recall and reports of how they had weighted the information differed significantly from control subjects' and were significantly less likely to correspond to the ratings of faculty members of how this information ought to be used. In addition, these subjects were less likely to register for and somewhat less likely to remain in courses that students who had taken the courses previously said were the best courses. Thus, regardless of whether the opinions of faculty members or students' peers (those who had previously taken the courses) were used as the criteria of an optimal choice, subjects in the rate all condition appeared to have made less optimal choices than control subjects. We predicted that the rate all manipulation would change subjects' choices by moderating their evaluations, so that the courses appeared more similar to each other. We found two pieces of evidence in support of this prediction. Both the range in their ratings of how likely they were to take the courses and the range in their ratings of how much they were influenced by the different information about the courses were significantly smaller than the ranges in the other two conditions.

Asking subjects to analyze the reasons for their evaluations of the courses also caused them to weight the course information in a less optimal way and to make less optimal choices. The effects of this manipulation, however, were not as strong as the effects of the rate all manipulation. On some measures, subjects who analyzed reasons responded similarly to control subjects, such as on their reports of how the different kinds of course

information influenced their decisions. On those measures that were most objective and consequential, however, our predictions were confirmed. For example, subjects in the reasons condition were significantly less likely than control subjects to preregister for and enroll in courses that had received high course evaluations (see Table 4). In addition, the correspondence between their recall of the course information and faculty members' ratings of this information was significantly lower than it was for control subjects (see Table 2).

As predicted, analyzing reasons did not make the courses seem more similar to subjects. In fact, the range in their ratings of the courses was significantly larger than it was in the control condition. Nor did analyzing reasons lower the range in their ratings of how much they were influenced by the different kinds of information about the courses. Thus, subjects in the reasons condition seemed to have had little difficulty in forming an opinion about which courses they liked and how the course information influenced them; it is just that their opinions differed from control subjects' (at least as assessed by their recall of the course information and the courses for which they registered and in which they were enrolled). These results are consistent with our hypothesis that when people analyze their reasons, they often change their criteria by focusing on attributes that seem like plausible reasons for liking or disliking the attitude object, but that in fact have not been heavily weighted before. Similarly, they dismiss attributes that seem like implausible reasons, but that in fact have been weighted heavily before. As a result, people change their mind about how they feel.

Despite this support for our predictions, we should not overlook the inconsistent effects of the reasons manipulation in Study 2 (e.g., the failure of this manipulation to influence subjects' reported likelihood of taking the courses). We offer the following, speculative explanation for these inconsistent findings. Both Wilson, Dunn, Kraft, and Lisle (1989) and Millar and Tesser (1986a) suggested that analyzing reasons is most likely to change attitudes that have a large affective component, because people are less likely to know the actual causes of these attitudes and because analyzing reasons is likely to emphasize cognitions and obscure the affect (the Millar & Tesser (1986a) explanation). People's attitudes toward college courses may have less of an affective component than their attitudes toward food items (e.g., strawberry jams), explaining why the effects were less consistent in Study 2. In addition, analyzing reasons may have a greater effect when the different dimensions of the stimuli are ill-defined, because this increases the likelihood that people will overlook factors that initially influenced their judgments. Consistent with this view, the criteria used to evaluate the courses in Study 2 were much more explicit than were the criteria in Study 1. That is, in Study 2, we gave subjects a list of all the relevant attributes of the different courses, whereas in Study 1, subjects had to define the set of relevant attributes themselves (e.g., whether to consider the color or consistency of the jams). Clearly, further research is needed to verify these speculations.

Finally, we should mention a possible alternative explanation for the effects of the introspection manipulations. The manipulations may have caused people to attend less to the information about the courses, because they were concentrating on why they felt the way they did. According to this argument, any

intervention that distracts people from the information about the alternatives would have similar deleterious effects to our introspection manipulations. The results of our recall measure, however, reduce the plausibility of this interpretation. If subjects in the introspection conditions were distracted, they should have recalled less information about the courses than did control subjects; in fact, there were no significant differences between conditions in the amount of information they recalled—only, as predicted, in the kinds of information they recalled (see Table 2).

General Discussion

Previous studies demonstrated that thinking about why we feel the way we do could change our attitudes (Wilson, 1990; Wilson, Dunn, Kraft, & Lisle, 1989). It has not been clear, however, whether the direction of this change is beneficial, detrimental, or neutral. The present studies demonstrated that analyzing reasons can lead to preferences and decisions that correspond less with expert opinion. This result, taken together with Wilson et al.'s (1990) finding that analyzing reasons reduces people's satisfaction with their choices, suggests that it may not always be a good idea to analyze the reasons for our preferences too carefully. In the present studies, analyzing reasons focused subjects' attention on characteristics of the stimuli that were, according to expert opinion, nonoptimal and caused them to use these characteristics to form preferences that were also nonoptimal. Nor may it be wise to analyze the effects of every attribute of every alternative. Evaluating multiple attributes led to nonoptimal preferences in Study 2 by moderating people's evaluations, so that the college courses seemed more equivalent than they did to subjects in the other conditions.

We do not mean to imply that the two kinds of introspection we examined will always lead to nonoptimal choices, and we certainly do not suggest that people studiously avoid all reflection before making decisions. Such a conclusion would be unwarranted for several reasons. First, we used stimuli in the present studies that were evaluated fairly optimally by control subjects, who were not instructed to reflect about the alternatives. That is, the evaluations and choices of control subjects in both studies corresponded fairly well with the experts' ratings. If people start out with feelings or preferences that are nonoptimal, the change that often results from introspection may be in a positive direction. Consistent with this possibility, Tesser, Leone, and Clary (1978) found that when people who experienced speech anxiety were asked to think about why they felt anxious, their anxiety was reduced.

Second, some people might be more likely to know why they feel the way they do about an attitude object and thus will be less likely to be misled by thinking about their reasons. Consistent with this hypothesis, Wilson, Kraft, and Dunn (1989) found that people who were knowledgeable about the attitude object and thus more likely to have attitudes that were based on objective, easily verbalizable attributes of it were relatively immune to the effects of thinking about reasons. Finally, in our studies, people were asked to reflect for a relatively brief amount of time. A more intensive, in-depth analysis, such as that advocated by Janis and Mann (1977), may have very differ-

ent effects on the quality of people's decisions (see, for example, Mann, 1972).

We have just begun to explore the conditions under which people should and should not reflect about the reasons for their preferences, thus to make broad claims about the dangers of introspection would be inappropriate (or at least premature). Perhaps the best conclusion at this point is a variation of Socrates' oft-quoted statement that the "unexamined life is not worth living." We suggest that, at least at times, the unexamined choice is worth making.

References

- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46, 610-620.
- Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology*, 12, 473-498.
- Goodman, N. G. (Ed.). (1945). *A Benjamin Franklin reader*. New York: Crowell.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, 93, 258-268.
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Judd, C. M., & Lusk, C. M. (1984). Knowledge structures and evaluative judgments: Effects of structural variables on judgmental extremity. *Journal of Personality and Social Psychology*, 46, 1193-1207.
- Keeney, R. L. (1977). The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance*, 19, 267-310.
- Kimble, G. A., & Perlmuter, L. C. (1970). The problem of volition. *Psychological Review*, 77, 361-384.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Langer, E. J., & Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology*, 37, 2014-2024.
- Linville, P. W. (1982). The complexity-extremity effect and age-based stereotyping. *Journal of Personality and Social Psychology*, 42, 193-211.
- Mann, L. (1972). Use of a "balance sheet" procedure to improve the quality of personal decision making: A field experiment with college applicants. *Journal of Vocational Behavior*, 2, 291-300.
- Millar, M. G., & Tesser, A. (1986a). Effects of affective and cognitive focus on the attitude-behavior relationship. *Journal of Personality and Social Psychology*, 51, 270-276.
- Millar, M. G., & Tesser, A. (1986b). Thought-induced attitude change: The effects of schema structure and commitment. *Journal of Personality and Social Psychology*, 51, 259-269.
- Millar, M. G., & Tesser, A. (1989). The effects of affective-cognitive consistency and thought on the attitude-behavior relation. *Journal of Experimental Social Psychology*, 25, 189-202.
- Morrow, J., & Nolan-Hoeksema, S. (1990). Effects of responses to depression on the remediation of depressive affect. *Journal of Personality and Social Psychology*, 58, 519-527.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Raiffa, H. (1968). *Decision analysis*. Reading, MA: Addison-Wesley.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22, 36-71.
- Slovic, P. (1982). Toward understanding and improving decisions. In W. C. Howell & E. A. Fleishman (Eds.), *Information processing and decision making* (pp. 157-183). Hillsdale, NJ: Erlbaum.
- Strawberry jams and preserves. (1985, August). *Consumer Reports*, pp. 487-489.
- Tesser, A., Leone, C., & Clary, G. (1978). Affect control: Process constraints versus catharsis. *Cognitive Therapy and Research*, 2, 265-274.
- Wilson, T. D. (1990). Self-persuasion via self-reflection. In J. M. Olson & M. P. Zanna (Eds.), *Self-inference processes: The Ontario Symposium* (Vol. 6, pp. 43-67). Hillsdale, NJ: Erlbaum.
- Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology*, 47, 5-16.
- Wilson, T. D., Dunn, D. S., Kraft, D., & Listle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). San Diego, CA: Academic Press.
- Wilson, T. D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology*, 25, 379-400.
- Wilson, T. D., Lisle, D. J., & Schooler, J. (1990). *Some undesirable effects of self-reflection*. Unpublished manuscript, University of Virginia, Department of Psychology, Charlottesville.
- Wilson, T. D., Hodges, S. D., & Pollack, S. (1990). *Effects of explaining attitudes on survey responses: The moderating effects of attitude accessibility*. Unpublished manuscript, University of Virginia, Department of Psychology, Charlottesville.

Received January 5, 1990
 Revision received July 9, 1990
 Accepted July 12, 1990 ■